AN ANALYSIS OF THE WEIGHTED LEAST SQUARES TECHNIQUE AS A METHOD FOR THE CONSTRUCTION OF TREE VOLUME TABLES

by

DONALD DEANE MUNRO

A THESIS

submitted to

OREGON STATE UNIVERSITY

in partial fulfillment of
the requirements for the
degree of

MASTER OF SCIENCE

June 1964

AN ABSTRACT OF THE THESIS OF

Donald Deane Munro for the M. S. (Degree) in Forest Management. (Major)

Date thesis is presented

Title AN ANALYSIS OF THE WEIGHTED LEAST SQUARES TECHNIQUE AS A METHOD FOR THE CONSTRUCTION OF TREE VOLUME TABLES.

Abstract approved

(Major professor)

The purpose of this thesis was to investigate the use of weights in least squares regression volume table construction and thereby to determine the importance of the assumption of homogeniety of tree volume variance. Several weighted and unweighted linear regression equations were investigated using data from 340 Douglas-fir *Pseudotsuga menziesii* (Mirb.) Franco trees from the interior of British Columbia, Canada. The results of the analyses showed that: a) the variance of tree volume for large trees is up to 50 times greater than the variance for small trees; b) the variance of tree volume is directly related to the square of the quantity FD^2H and also the square of the quantity D^2H; c) erroneous statistical conclusions may be reached if statistical tests are carried out for unweighted regression solutions for tree volume because of the large differences

in volume variance; d) the use of weighted least squares regression analysis improved the volume estimate precision for all equation forms tested.

ACKNOWLEDGEMENTS

The author wishes to express his gratitude to the Vancouver Forest Products Laboratory of the Canada Department of Forestry for permission to use the data; to Antal Kozak and the University of British Columbia Computing Centre for assistance and cooperation; to the School of Forestry at Oregon State University and the Lewis W. and Maud Hill Foundation for generous financial assistance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

AN ANALYSIS OF THE WEIGHTED LEAST SQUARES TECHNIQUE AS A METHOD FOR THE CONSTRUCTION OF TREE VOLUME TABLES

INTRODUCTION

Assumptions of the Least Squares Technique

The least squares method of regression analysis has long been used in tree volume table construction. In many instances, however, the following mathematical assumptions have been considered unimportant:

a) normality: the conditional distribution of tree volume on the independent variables must be normally distributed.

b) homogeniety of variance: the variance of tree volume must be homogeneous throughout the range of the independent variables.

c) randomness: sample trees must be randomly selected.

Usually none of these assumptions are satisfied in conventional tree volume table construction. The conditional distribution of tree volume is often skewed; the variance of tree volume is not homogeneous; and the sample trees are selected through a selective or systematic sampling design.

If the assumptions of conditional normality and randomness in sampling are not fulfilled, the effect is not serious in the estimation of the regression coefficients. The probability level of the significance tests and the confidence limits will be affected, however (6, p. 2).

The assumption of homogeniety of tree volume variance is probably the most important in tree volume table construction. If the variance of the dependent variable is larger for some values of the independent variable than for others, then those values with high variance will have a disproportionate effect on the least squares estimation of the regression coefficients. In addition, the confidence limits and the tests of significance will not be accurate. Tree volume variance is commonly much larger for bigger trees than for smaller trees (6, p. 2, 4; 9; 10; 13).

Purpose and Scope

The purpose of this thesis is to investigate the use of weights in least squares regression volume table construction and thereby to determine the importance of the assumption of homogeniety of tree volume variance. The weighted and unweighted least squares regression solutions of four common tree volume equations will be compared. The importance of the variance homogeniety assumption will be evaluated on the basis of the results obtained from these equations using Douglas-fir *Pseudotsuga menziesii* (Mirb.) Franco tree volumes from the interior of the province of British Columbia, Canada.

No attempt will be made to correct for lack of conditional normality or for non-random sampling methods. Neither will any intensive investigation be undertaken to determine the best form of the tree volume equation. This has been done by others (11; 17, p. 93).

THE BASIC DATA

Data Collection

During the summer of 1962, the author was employed as a forest assistant by the utilization section of the Forest Products Laboratory of the Canada Department of Forestry at Vancouver, British Columbia. The field work consisted of collecting data on mature Douglas-fir *Pseudotsuga menziesii* (Mirb.) Franco from the interior of the province of British Columbia for the purpose of establishing a series of tree and log grading rules. Many tree variables were measured, including those necessary for the determination of total gross tree volume in cubic feet. In April 1963, the Canada Department of Forestry made available to the author the data for 340 of these trees. The areas of collection (Table I) were centered near Vernon in the Okanagan Valley and Williams Lake in the central interior of the province.

Measurements of the following variables were available on electronic data processing punch cards: total tree height in feet, diameter at breast height outside bark in inches, diameter inside bark in inches at approximately 16-foot intervals throughout the merchantable length of the tree, stump height in inches, stump diameter in inches, merchantable length in feet, and gross merchantable volume of the tree in cubic feet.

Table I

Source and Number of Sample Trees

Locality	Number of Trees
Vernon Area:	
Lumby	18
Monte Lake	43
Falkland	30
Oyama	29
Silver Creek	19
Enderby	10
Pinaus Lake	28
Williams Lake Area:	
Six Mile	52
Horse Fly	51
Joes Lake	60
Total	340

Data Summarization

All of the data summarization was carried out during the summer of 1963. Use of the I.B.M. 1620 Electronic Computer at the University of British Columbia Computing Centre was made whenever possible.

Because the data were not collected for the purpose of obtaining total gross tree volumes in cubic feet, considerable summarization was required. The values of the variables previously named were not available on a single set of punch cards. A computer program was written to merge the values of these variables onto one set of punch cards. Following merging, the total gross tree volumes in cubic feet were calculated. This necessitated the computation of stump and top volumes and their addition to the gross merchantable volume already available on the punch cards.

Stump volumes were calculated on the basis of a cylinder having a diameter equal to the top diameter of the stump and a height equal to the height of the stump. Top form was assumed to be between paraboloid and conic frustrums. Top volumes were therefore calculated on the basis of 0. 4 times the basal area times the length of the top. A computer program was written to perform the above computations.

Diameter measurements inside bark at 16 feet and 32 feet were available only to the nearest inch. Thus Girard form class determinations were not as precise as they could have been if diameter measurements to the nearest tenth of an inch had been available.

The statistics of the summarized data are presented in Table II.

Table II

Statistics of Basic Data

Variable Name	Mean	Standard Deviation
Total tree volume (cu. ft.)	81.7	66.5
DBH (in.)	21.2	6.4
Total tree height (ft.)	94.1	24.2
Girard Form Class (16 ft.)	.72	.07
Girard Form Class (32 ft.)	.62	.09

TEST EQUATIONS

Equation Form

In order to compare results of the weighted and unweighted least squares analysis, it was necessary to determine the form of the equation which suited the data. Several authors have investigated this problem (11; 17, p. 93).

For the purposes of this analysis, several alternative equation forms were desired for comparison. To facilitate the selection of these equations, the unweighted least squares multiple linear regression program available at the University of British Columbia Computing Centre was used. The five basic variables (Table II) were assessed singly and in combination with a total of ten transformations and combinations. In addition, several generally accepted tree volume equations (17, p. 97) were solved (Table III). The unweighted least squares solutions of these equations were compared according to the residual variance and the coefficients of determination. It was concluded that equations 1, 2, 3, and 5 in Table III could all be considered to fit the data reasonably well. It was therefore decided to use these four equations as test equations throughout the remainder of the investigation.

Table III

Tree Volume Equations*

Australian	$V = a + bD^2 + cH + dD^2H$
Schumacher	$\log V = a(\log D) + b(\log H) + \log (c)$
Combined Variable	$V = a + bD^2H$
Combined Variable Form Class	$V = a + bF + cD^2H + dFD^2H$
Short Cut Form Class	$V = a + bFD^2H$

* In the above equations V is volume in cubic feet, D is diameter at breast height outside bark, H is total height, and F is Girard Form Class. Lower case letters are coefficients which vary according to the formula and data used.

Origin of Test Equations

Australian. T. N. Stoate (19) experimenting with data from the South Australian Department of Forestry derived the equation form known as the Australian Equation. After logarithmic expressions proved unsuccessful in the estimation of volume of *Pinus radiata*, the joint function of basal area and height was added as a third independent variable. Stoate's tests of this equation were confined to trees in the seven to ten inch diameter class. Spurr (17, p. 103) later tested this equation and found it useful for a wide range of diameters and species.

Schumacher. Schumacher and Hall (15) were the first American foresters to investigate the mathematical relationship between tree volume and diameter and height. They reasoned that because tree form was nearly always correlated with height and diameter, that volume of trees could not be expressed as a function of diameter squared and height. They concluded that volume increased with certain powers of diameter and height other than two and one respectively. Adopting this assumption, the tree volume equation can then be written:

$$V = D^{a} \cdot H^{b} \cdot C$$

where V is volume, D is diameter at breast height outside bark, C is a constant and "a" and "b" are coefficients. This may be expressed logarithmically in the form:

$$\log V = a(\log D) + b(\log H) + \log (c)$$

which is known as Schumacher's tree volume equation. An equation of this type was first applied to 264 yellow poplar trees with excellent results. Since that time it has become widely accepted as a standard tree volume equation throughout North America.

Combined Variable. Spurr (17, p. 111) first used the combined variable formula as a solution for stand volume estimation and later applied it to the estimation of individual tree volumes. He is generally credited with the derivation of the formula. In actual fact, however, Stoate first suggested the use of this equation for tree volume estimation. He said (19):

> It will be seen that almost as good a fit as any other is obtained by using the joint function only (basal area times height), dropping both independent variables basal area and height. This is of interest in that except for the deduction of a small constant, this is the time-honoured method of the forester, the reduction of the product of basal area and height by a form factor.

Short Cut Form Class. Spurr (17, p. 96) derived this formula as a short form of the combined variable form class solution in a manner analagous to the derivation of the combined variable solution. He found it to be one of the simplest and most satisfactory of total cubic foot tree volume equations tested.

Theory of Weighted Least Squares

A minimum of the theory and calculation procedures necessary to carry out weighted least squares regression solutions will be presented. No attempt will be made to offer detailed proof of the least squares theorem and the application of weights to it. The reader is referred to (1, p. 186; 7) for this information.

As stated in the introduction, homogeniety of variance of the dependent variable is a basic assumption in least squares estimation. Because the least squares technique minimizes the sum of squares of deviations from the regression line, those classes of the dependent variable which have excessively high variance (and thus excessively high deviations from the regression line) will have a disproportionate effect on the estimation of the regression coefficients. One way to remedy the situation is to weight the dependent variable in such a manner that the variance is made homogeneous throughout the range of the independent variable. The simplest way of accomplishing this is to multiply each variable in the equation by the inverse of the variance of the dependent variable (21, p. 19.)

Consider the observation equation of the form:

$$y_i = \beta_0 + \beta_1 x_i$$

where: y_i is an independent random variable

x_i is a fixed variate

β_0, β_1 are regression coefficients.

Weighting this equation by the inverse of the variance of the dependent variable $\sigma^2 y_i$, results in the equation:

$$\frac{y_i}{\sigma^2 y_i} = \frac{\beta_0}{\sigma^2 y_i} + \frac{\beta_1 x_i}{\sigma^2 y_i}$$

The normal equations for a series of such observation equations can be written as follows:

$$\Sigma \left[\frac{y_i}{\sigma^2 y_i} \right] = \beta_0 \, \Sigma \left[\frac{1}{\sigma^2 y_i} \right] + \beta_1 \, \Sigma \left[\frac{x_i}{\sigma^2 y_i} \right]$$

$$\Sigma \left[\frac{y_i}{\sigma^2 y_i} \cdot \frac{x_i}{\sigma^2 y_i} \right] = \beta_0 \, \Sigma \left[\frac{1}{\sigma^2 y_i} \cdot \frac{x_i}{\sigma^2 y_i} \right] + \beta_1 \, \Sigma \left[\frac{x_i}{\sigma^2 y_i} \right]^2$$

Solution of the above normal equations will provide an unbiased, efficient estimate of the regression coefficients (21, p. 19).

Now if it is possible to relate the variance of the dependent variable to the independent variable in some manner, solutions of the equations can be further simplified. For example, if $\sigma^2 y_i$ can be shown to be proportional to x_i^2 , then the variables can be weighted by the inverse of x_i^2 rather than the inverse of $\sigma^2 y_i$. Thus:

$$\frac{y_i}{\sigma^2 y_i} = \frac{\beta_0}{\sigma^2 y_i} + \frac{\beta_1 x_i}{\sigma^2 y_i}$$

can be written as:

$$\frac{y_i}{x_i^2} = \frac{\beta_0}{x_i^2} + \frac{\beta_1 x_i}{x_i^2}$$

which reduces to:

$$\frac{y_i}{x_i} = \frac{\beta_0}{x_i} + \beta_1$$

and the normal equations are:

$$\Sigma \left[\frac{y_i}{x_i}\right] = \beta_0 \Sigma \left[\frac{1}{x_i}\right] + n\beta_1$$

$$\Sigma \left[\frac{y_i}{x_i} \cdot \frac{1}{x_i}\right] = \beta_0 \Sigma \left[\frac{1}{x_i}\right]^2 + \beta_1 \Sigma \left[\frac{1}{x_i}\right]$$

The least squares solutions of these normal equations is equivalent to the weighted least squares solution of the equation form:

$$y_i = \beta_0 + \beta_1 x_i \quad .$$

The general theory outlined above can be expanded and applied to multiple regression solutions if desired.

DATA ANALYSES

Tree Volume Variance Analyses

Preliminary graphical examination of the data indicated that the volume variance was not homogeneous throughout the range of the independent variables. A series of computer programs was written to sort the dependent variable volume into classes according to the various independent variables. Average volumes and volume variances for successive classes of each independent variable were calculated. Graphical analyses were then used in an attempt to find a relationship between volume variance and one or more of the independent variables. The relationship of volume variance to dbh, dbh squared, height, height squared, Girard form class, D^2H, D^2H squared and FD^2H squared was examined in this manner. It was concluded that the variance of volume was linearly associated with only two of the variables examined--D^2H squared and FD^2H squared (Figures 1, 2). This finding substantiates the results of (6, p. 2; 13).

In order to demonstrate the application of the weighted least squares regression method and to assess its usefulness in tree volume estimation, weighted and unweighted least squares solutions were solved for the test equations described on pages 9-11.

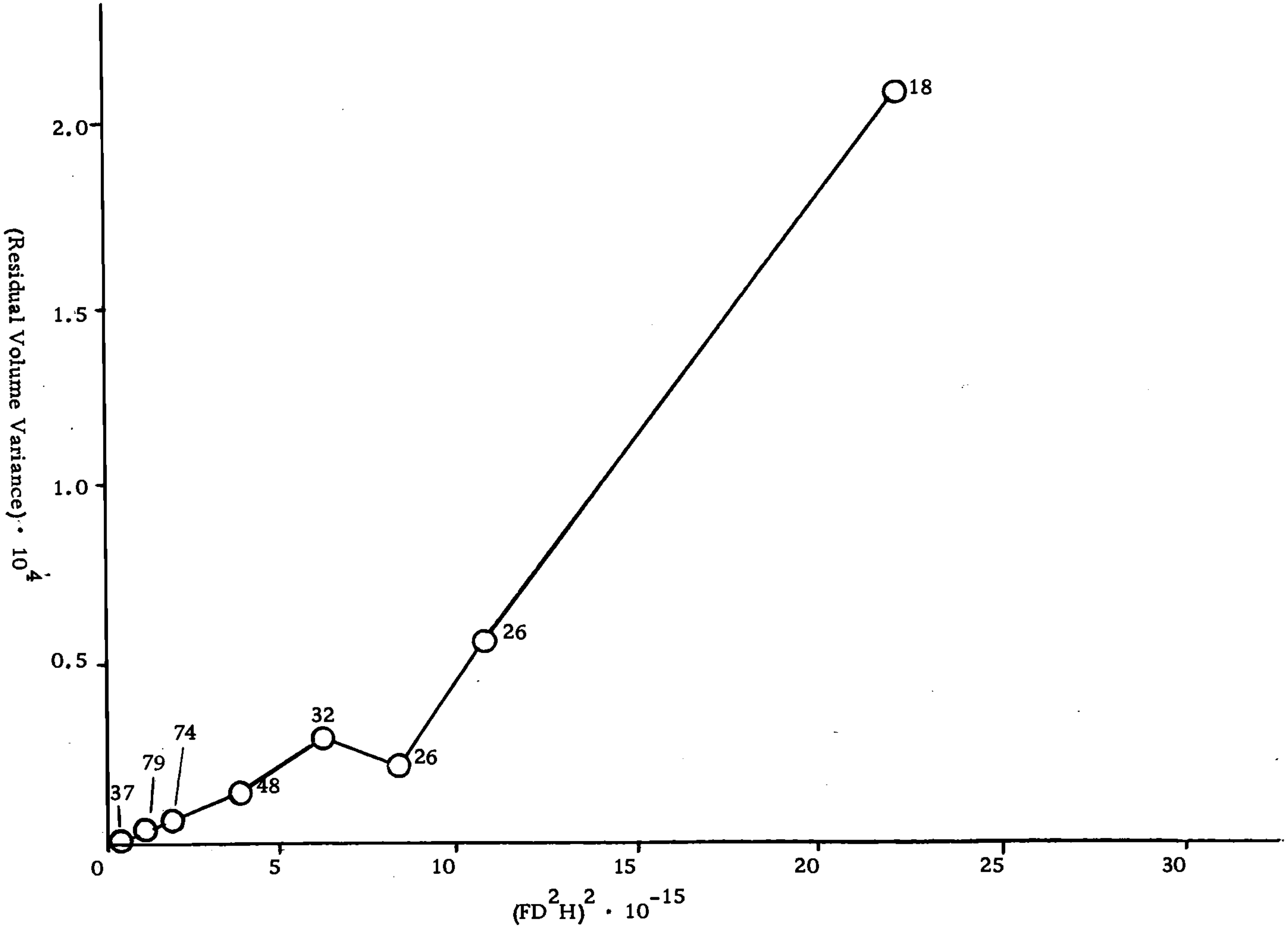


Figure 1. Graph showing the relationship between residual volume variance and $(FD^2H)^2$.

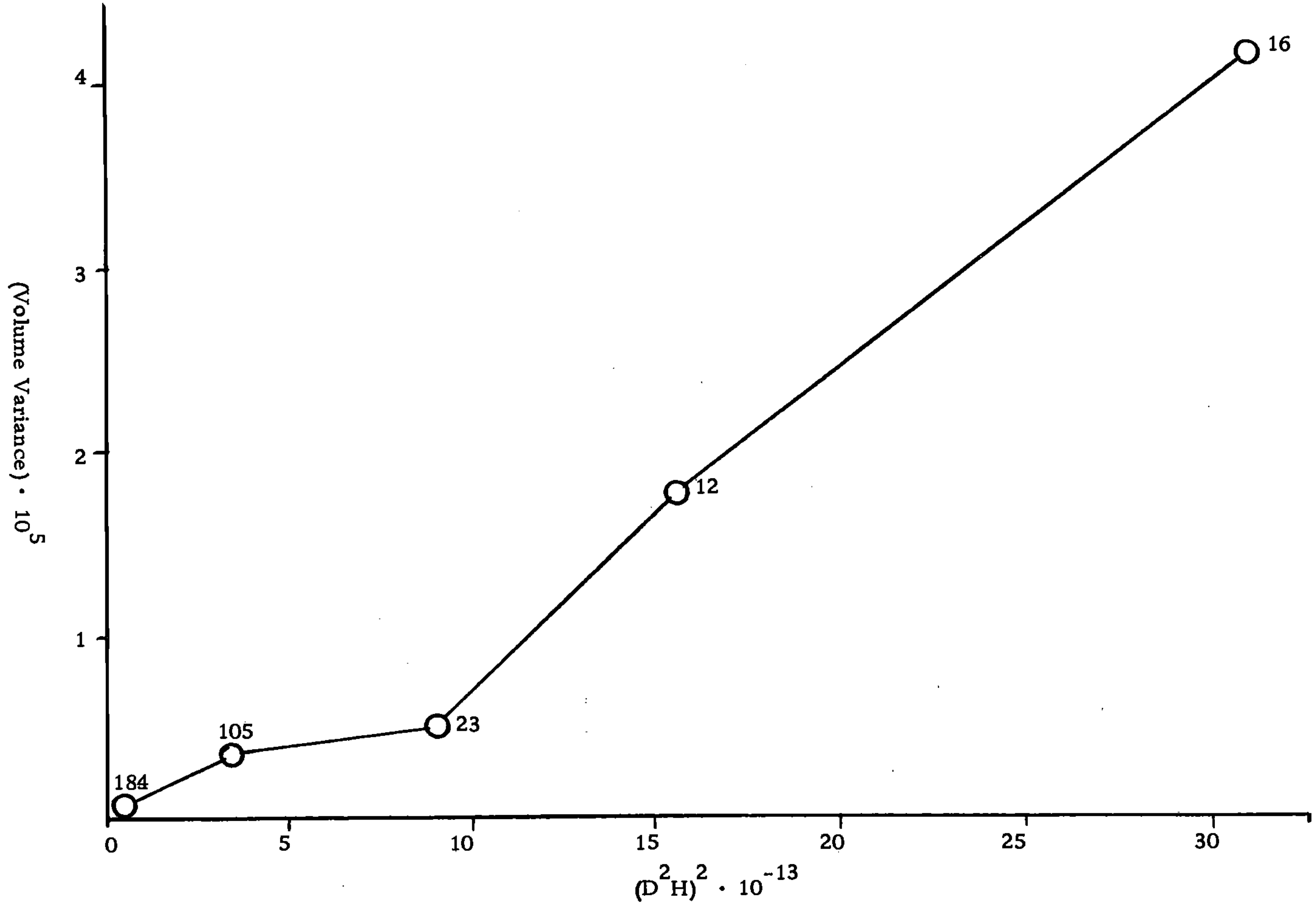


Figure 2. Graph showing the relationship between volume variance and $(D^2H)^2$.

using $\frac{1}{(D^2H)^2}$ and $\frac{1}{(FD^2H)^2}$ as weights.

Short Cut Form Class Equation

The unweighted least squares solution for this equation was:

$$V = 43.28270 + 0.2169489\ FD^2H \cdot 10^{-5}$$

where V = total tree volume in cubic feet times ten

F = Girard form class at 16 feet (whole number)

D = diameter at breast height outside bark in tenths of inches

H = total tree height in feet.

There is no commonly known way to calculate confidence intervals or statistical tests of this regression solution due to the lack of homogeniety of volume variance.

The least squares solution of the weighted variables was of the form:

$$\frac{V}{FD^2H \cdot 10^{-5}} = \frac{\beta_0}{FD^2H \cdot 10^{-5}} + \beta_1$$

It gave the solution:

$$\frac{V}{FD^2H \cdot 10^{-5}} = \frac{26.22715}{FD^2H \cdot 10^{-5}} + 0.22240$$

which when multiplied through by $FD^2H \cdot 10^{-5}$ gives the equation:

$$V = 26.22715 + 0.22240\ FD^2H \cdot 10^{-5}\ .$$

Valid significance tests and confidence intervals can be calculated for this solution. This equation turned out to be the most efficient volume estimator. To show how some useful statistical tests can be computed for the weighted least squares solutions, calculation procedures for several statistics will be illustrated for this equation only. Some of the statistics illustrated are not calculated for the other test equations.

Confidence limits for the estimated mean volume for any given value of the independent variable can be calculated with the following formula (12, p. 287):

$$\text{C. L.} = t\sqrt{\text{MSE}}\sqrt{\frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\text{SSX}}}$$

where t = percentage point of student's t distribution.

MSE = mean square residual variance from the analysis of variance.

n = the number of observations

X_0 = selected value of X_i for which the C. L. is desired

$\overline{X}$ = mean of all X's

SSX = the corrected sum of squares of the X's.

Using values from the weighted least squares solution, the 95 percent confidence limit for the estimated mean $V/FD^2H \cdot 10^{-5}$ for the mean value of $1/FD^2H \cdot 10^{-5}$ would be:

$$1.96\sqrt{.000943}\sqrt{\frac{1}{340} + \frac{(.00048 - .00048)^2}{.000048376}} = .0032636 \;.$$

It must be realized that this limit is applicable only to the transformed variables used in the weighted solution. If expressed as a percent of the estimated $V/FD^2H \cdot 10^{-5}$, however, it can then be applied to the actual values of volume. The confidence limit .0032636 is 1.389 percent of the estimated $V/FD^2H \cdot 10^{-5}$ where $1/FD^2H \cdot 10^{-5} = .00048$, and thus the confidence interval for V is ±1.389 percent of the estimated V where $FD^2H \cdot 10^{-5} = 1/.00048$. Confidence limits for various values of $FD^2H \cdot 10^{-5}$ are tabulated in Table IV.

A confidence limit for β_0, the regression parameter for slope in the weighted regression can be calculated from the formula (12, p. 282):

$$C.L. = t\sqrt{\frac{MSE}{SSX}}$$

where the symbols are defined on page 19. Using values from the weighted least squares solution, the 95 percent confidence interval for β_0 is:

$$26.22715 \pm 1.96\sqrt{\frac{.000943}{.000048376}} = 17.5737 \text{ to } 34.8805 \quad .$$

The confidence limit for β_1, the intercept in the weighted least squares solution, is identical to the percent confidence limit calculated for the estimated $V/FD^2H \cdot 10^{-5}$ at the mean value of the independent variable on page 19. This is 1.389 percent. Thus for the weighted least squares solution, the 95 percent confidence interval for β_1 is:

$$.22240 \pm 1.389 \text{ percent or } .21931 \text{ to } .22549 \quad .$$

Tolerance limits, the range in which a single estimate of volume for any given value of the independent variable is expected to lie, can be calculated with the formula (21, p. 99):

$$\text{T. L.} = t\sqrt{\text{MSE}}\sqrt{1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\text{SSX}}}$$

where symbols are defined on page 19. Using values from the weighted least squares solution, the 95 percent tolerance limit for the single estimated $V/FD^2H \cdot 10^{-5}$ for the mean value of $1/FD^2H \cdot 10^{-5}$ would be:

$$1.96\sqrt{.000943}\sqrt{1 + \frac{1}{340} + \frac{(.00048 - .00048)^2}{.000048376}} = .06028 \quad .$$

Expressed as a percentage of the estimated $V/FD^2H \cdot 10^{-5}$ where $1/FD^2H \cdot 10^{-5} = .00048$, this is a 25.65 percent. Tolerance limits for various values of $FD^2H \cdot 10^{-5}$ are tabulated in Table IV.

The weighted least squares regression line, and confidence and tolerance intervals are presented graphically in Figure 3. The significance of the regression is tested by the analysis of variance in the conventional manner in Table V.

Australian Equation

The unweighted least squares solution of this equation form was:

$$V = -183.7896 + .0013132D^2 + 3.037306H + 1.30492D^2H \cdot 10^{-4} \quad .$$

Table IV

Confidence and Tolerance Limits for Weighted Short Cut Form Class Equation

$FD^2H \cdot 10^{-5}$	Estimated Volume	95% Confidence Limit - %	95% Tolerance Limit - %
833	21.1	2.77	23.87
1000	24.8	2.24	24.31
1250	30.4	1.76	24.79
1667	39.6	1.44	25.32
* 2080	49.0	1.39	25.65
2500	58.2	1.43	25.88
5000	114	1.78	26.50
10000	225	2.19	26.84
20000	447	2.21	26.99

* mean value

Table V

Analysis of Variance for Weighted Regression of Volume on FD^2H

Source	D.F.	Sum Squares	Mean Square	F
Regression	1	0.03788	0.03788	40.17*
Residual	338	0.31879	0.00094	
Total	339	0.35667		

* significant at the .01 percent level

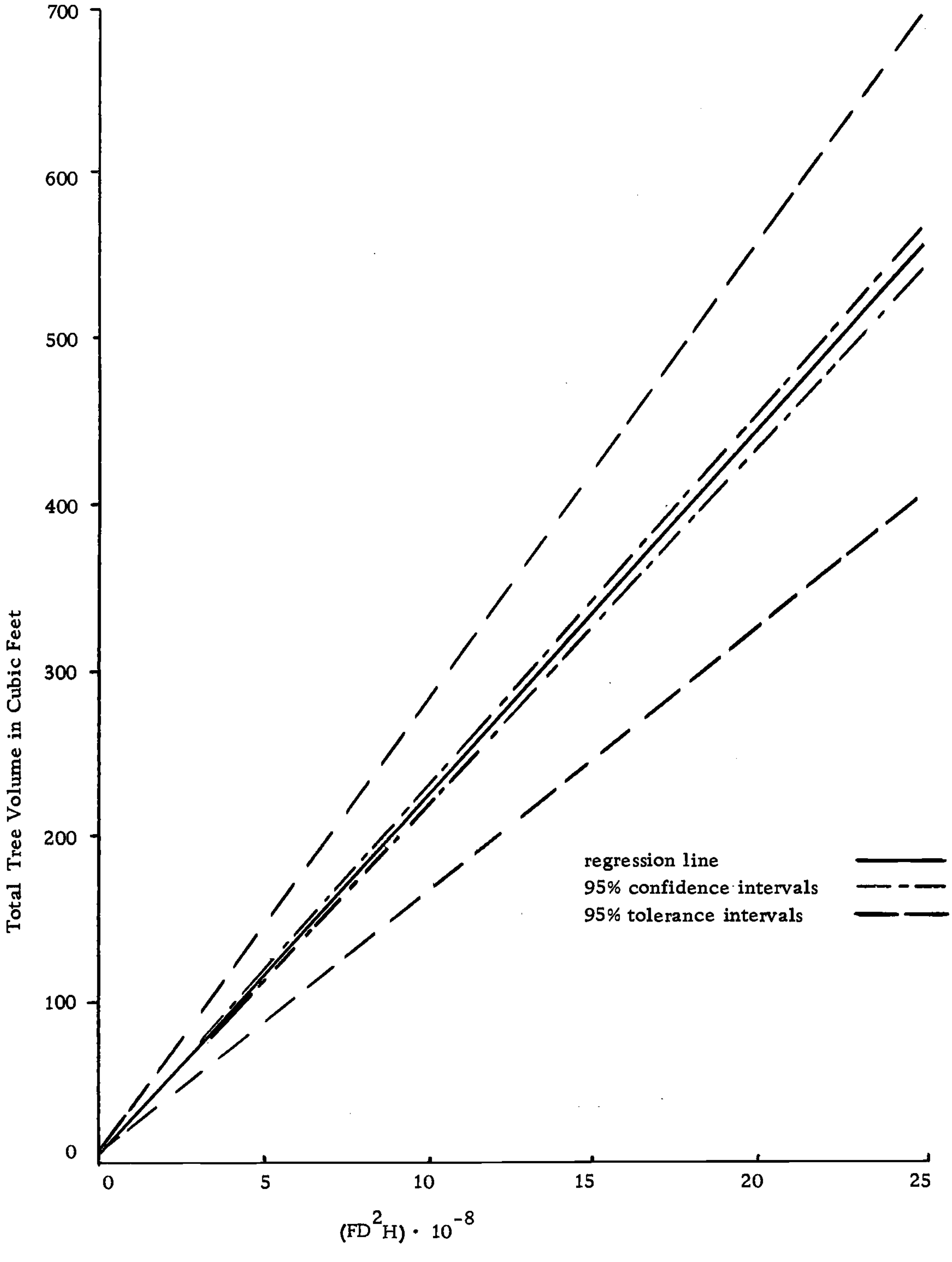


Figure 3. Regression line, confidence intervals, and tolerance intervals for weighted short cut form class regression solution

There is no commonly known way to calculate confidence intervals or statistical tests of this solution due to the lack of homogeniety of variance.

The least squares solution of the weighted variables is of the form:

$$\frac{V}{D^2H \cdot 10^{-4}} = \frac{\beta_0}{D^2H \cdot 10^{-4}} + \frac{\beta_1 D^2}{D^2H \cdot 10^{-4}} + \frac{\beta_2 H}{D^2H \cdot 10^{-4}} + \beta_3$$

It gives the solution:

$$\frac{V}{D^2H \cdot 10^{-4}} = \frac{29.57737}{D^2H \cdot 10^{-4}} + \frac{.00210D^2}{D^2H \cdot 10^{-4}} + \frac{.40744H}{D^2H \cdot 10^{-4}} + 1.7202$$

which when multiplied through by $D^2H \cdot 10^{-4}$ gives the solution:

$$V = 29.57737 + .00210D^2 + .40744H + 1.7202D^2H \cdot 10^{-4}$$

Valid significance tests and confidence intervals for this solution can be calculated by simple expansion of the formulae illustrated for the short cut formula. The analysis of variance is presented in Table VII. The standard error of the volume estimate for the mean values of the independent variables is presented in Table VI.

Combined Variable Equation

The unweighted least squares solution of this equation form is:

$$V = 63.29641 + 1.512046\, D^2H \cdot 10^{-4}$$

There is no commonly known way to calculate confidence intervals or statistical tests of this solution due to the lack of homogeniety of volume variance.

Table VI

Standard Error of Estimates of Single Tree Volumes for Test Equations

Equation	Standard Error (% of mean tree volume)
Short cut form class	
unweighted	14.1
weighted	13.0
Australian	
unweighted	20.4
weighted	17.2
Combined Variable	
unweighted	21.2
weighted	17.6
Schumacher	
logarithmic	20.7

Table VII

Analysis of Variance for Weighted Regression of Volume on D^2, H, and D^2H

Source	D.F.	Sum Squares	Mean Square	F
Regression	3	0.04661	0.015537	18.253*
Residual	336	0.28600	0.000852	
Total	339	0.33261		

* significant at the .01 percent level

The least squares solution of the weighted variables is of the form:

$$\frac{V}{D^2H \cdot 10^{-4}} = \frac{\beta_0}{D^2H \cdot 10^{-4}} + \beta_1$$

which gives the solution:

$$\frac{V}{D^2H \cdot 10^{-4}} = \frac{31.98409}{D^2H \cdot 10^{-4}} + 1.5875$$

which when multiplied through by $D^2H \cdot 10^{-4}$ is equal to:

$$V = 31.98409 + 1.5875\ D^2H \cdot 10^{-4}$$

Valid significance tests and confidence limits can be calculated for this solution. The analysis of variance is presented in Table VIII. The standard error of the volume estimate for the mean value of the independent variable is presented in Table VI.

Schumacher's Equation

One way of minimizing the problem of non-homogeniety of variance is to use the logarithms of tree volume (6, p. 3). Schumacher's equation is an example of the use of logarithms in the estimation of tree volume. The main disadvantage of this method is that the estimation of the arithmetic mean is replaced by the estimation of the geometric mean. Because the arithmetic mean is always larger than the geometric mean this equation always underestimates volume (6, p. 3). A second disadvantage of this method is that the standard error of estimate is based upon the variation of logarithms and it must be converted to natural values. Several formulae have been suggested to approximate this conversion (9; 17, p. 273).

Although the least squares analysis and statistical tests are valid for the logarithmic solution, the transformation of the results to natural values is approximate. Because of the wide use of Schumacher's equation it was used as a test equation in these analyses.

The solution for the data is:

$$\log V = -3.389054 + 1.71705 \log D + 1.134824 \log H \quad .$$

The analysis of variance in logarithmic form is presented in Table IX. The standard error of estimate, calculated directly from the predicted transformed values of the equation is presented in Table VI.

Table VIII

Analysis of Variance for Weighted Regression of Volume on D^2H

Source	D. F.	Sum Squares	Mean Square	F
Regression	1	0. 02884	0. 02884	32. 40*
Residual	338	0. 30377	0. 00089	
Total	339	0. 33261		

* significant at the . 01 percent level

Table IX

Analysis of Variance for Regression of Log Volume on Log D and Log H

Source	D. F.	Sum Squares	Mean Square	F
Regression	2	31. 557619	15. 7768	2209. 5*
Residual	337	2. 406651	. 0071414	
Total	339	33. 96427	0. 1001896	

* significant at the . 01 percent level

DISCUSSION

In assessing the accuracy and utility of tree volume equations several methods are available. The aggregate deviation (2) is the difference between the sum of the actual volumes and the sum of the estimated volumes expressed as a percentage of the latter. It is used extensively in checking the bias of volume tables constructed by graphical methods and alignment charts. It is generally not used in tables constructed by least square analysis, because the least squares technique itself insures that the aggregate deviation approximates zero.

The average deviation is simply the arithmetic sum of the absolute values of the differences between actual and estimated volumes expressed as a percentage of the sum of the estimated volumes. It is primarily an indication of the variability of the data used in the volume table construction. Although the average deviation has been used for many years as a check on volume table construction, its statistical significance cannot be assessed. In modern volume table construction this is an important criterion.

With the acceptance of the least squares technique, the standard error of the estimate, also known as the standard deviation from regression, has become the standard measure of the precision of a volume table equation (17, p. 75). This is the mean of the squares of the deviations of the actual from the estimated volumes. It may be calculated by squaring the deviations of each individual value from its estimate, but in least squares solutions it is commonly computed directly from the sums of squares and sums of products of the independent and dependent variables.

Another measure of statistical precision sometimes used in volume table construction is the correlation coefficient or the coefficient of determination, which is the square of the correlation coefficient. These values provide a relative measure of precision, but one which is strongly influenced by the assumption of bivariate normality. Bivariate normality is almost always absent in the basic data used in the construction of volume tables. For this reason the correlation coefficient has not been used as a test of precision in this analysis.

Standard errors of estimate in percent of mean tree volume were compared to assess the precision of the various test equations used in this analysis. These have been previously tabulated for all test solutions in Table VI on page 25. In each test equation, the standard error of estimate was considerably reduced when the weighted solution was used.

To show the effect of weighting on the least squares analysis, the percent deviations of the estimated volumes from the actual volumes were plotted over actual volume for weighted and unweighted solutions of each test equation (Appendices 1, 2, 3, 4). Also the estimated volumes for each test equation solution for various classes of actual volumes were arranged in tabular form for comparison purposes (Appendix 5).

Examination of the above graphs and tables indicates that the use of weights resulted in an improved fit in the smaller volume classes and a correction of the tendency in the unweighted solutions to overfit the equation to the larger volume classes.

In every case, the weighted least squares solution lowered the intercept and raised the slope compared to the unweighted solution. This suggests the possibility that the true regression line might be curved, although no curvilinearity could be detected in a graphical examination of the data. As mentioned previously, several authors have conducted intensive investigations into tree volume equation form, but only for unweighted solutions. Further research should be conducted to reconsider the best equation function in the light of a weighted analysis.

Some interesting results were observed when, in the course of the assessment of the precision of the test equations, analyses of variance were compared for the weighted and unweighted solutions. In every equation, the "F" statistics or variance ratios calculated for the unweighted solutions were much higher than for the comparable weighted solutions (Table X). The reason for the extremely high

"F" statistics in the unweighted solutions appears to be the lack of variance homogeniety. As illustrated in Figures 1 and 2, the variance for large tree volumes is up to 50 times greater than the variance for small tree volumes. The analyses of variance for the unweighted solutions are therefore not valid. Although in these test equations, the discrepancy in the "F" statistics was not large enough to result in erroneous statistical conclusions, they show clearly that it would be quite possible to obtain an indicated highly significant "F" statistic for an unweighted least squares solution, when in fact the true "F" statistic obtained from a weighted solution would show the regression to be not significant.

Table X

"F" Statistics or Variance Ratios for
Weighted and Unweighted Test Equation Solutions

	"F" Statistic	
Equation	Weighted Equation	Unweighted Equation
Short cut form class	40. 17	10, 938
Australian	18. 25	1, 673
Combined variable	32. 40	4, 634

Because the weighted short cut form class solution had the lowest standard error of estimate and fitted the data throughout the range better than any of the other test equations, a volume table for two-inch diameter classes and ten-foot height classes was constructed from this equation for the average form class of the data (Appendix 6).

CONCLUSION

The least squares method of regression analysis is valid only if the assumptions of conditional normality, homogeniety of variance, and randomness in sampling are fulfilled. In tree volume table construction the most important assumption is probably the homogeniety of variance. From the results of this study it is concluded that:

a) The variance of tree volume for large trees is up to 50 times greater than the variance of tree volume for small trees.

b) The variance of tree volume is directly related to the square of the quantity D^2H and also the square of the quantity FD^2H.

c) Erroneous statistical conclusions may be reached if statistical tests are carried out for unweighted regression solutions for tree volume because of the large differences in volume variance.

d) The use of weighted least squares regression analysis improved the volume estimate precision for all equation forms tested.

BIBLIOGRAPHY

1. Anderson, R. L. and T. A. Bancroft. Statistical theory in research. New York, McGraw-Hill, 1952. 399 p.

2. Bruce, Donald. A proposed standardization of the checking of volume tables. Journal of Forestry 18: 544-548. 1920.

3. Bruce, Donald and F. X. Schumacher. Forest mensuration. 3d ed. New York, McGraw-Hill, 1950. 483 p.

4. Buckman, R. E. Development and use of three stand volume equations for Minnesota. Journal of Forestry 59: 573-575. 1961.

5. Chapman, H. H. and W. H. Meyer. Forest mensuration. New York, McGraw-Hill, 1949. 522 p.

6. Cunia, T. Weighted least squares method and construction of volume tables, preliminary report. Paper submitted to Mensuration Group, Montreal, Quebec, September 17, 1962. 53 p. (Mimeographed)

7. David, F. N. and J. Neyman. Extension of the Markoff theorem on least squares. Statistical Research Memoirs 2: 105-116. 1938.

8. Dudzinski, M. L. Estimation of regression slope from tail regions with special reference to the volume line. Biometrics 16: 399-407. 1960.

9. Furnival, G. M. An index for comparing equations used in constructing volume tables. Forest Science 7: 337-341. 1961.

10. Gedney, D. R. and F. A. Johnson. Weighting factors for computing the relationship between tree volume and dbh in the Pacific Northwest. 1959. 5 p. (U. S. Dept. of Agriculture. Forest Service. Pacific Northwest Forest and Range Experiment Station, Portland. Research Note 174)

11. Golding, D. L. and O. F. Hall. Tests of precision of cubic-foot tree volume equations. Forestry Chronicle 37: 123-132. 1961.

12. Li, Jerome C. R. Introduction to statistical inference. Ann Arbor, Edwards Brothers, 1957. 568 p.

13. Palley, Marshall N. Board-foot volume tables for young growth coastal redwood. Berkeley, 1959. 3 p. (California. University, School of Forestry, California Forestry and Forest Products 11)

14. Schumacher, F. X. and R. A. Chapman. Sampling methods in forestry and range management. Durham, North Carolina, Duke University, 1942. 213 p.

15. Schumacher, F. X. and Francisco Dos Santos Hall. Logarithmic expression of timber tree volume. Journal of Agriculture Research 47: 719-734. 1933.

16. Smith, J. H. G. and John W. Ker. Timber volume depends on D^2H. British Columbia Lumberman 41: 28-30. Sept., 1957.

17. Spurr, S. H. Forest inventory. New York, Ronald, 1950. 476 p.

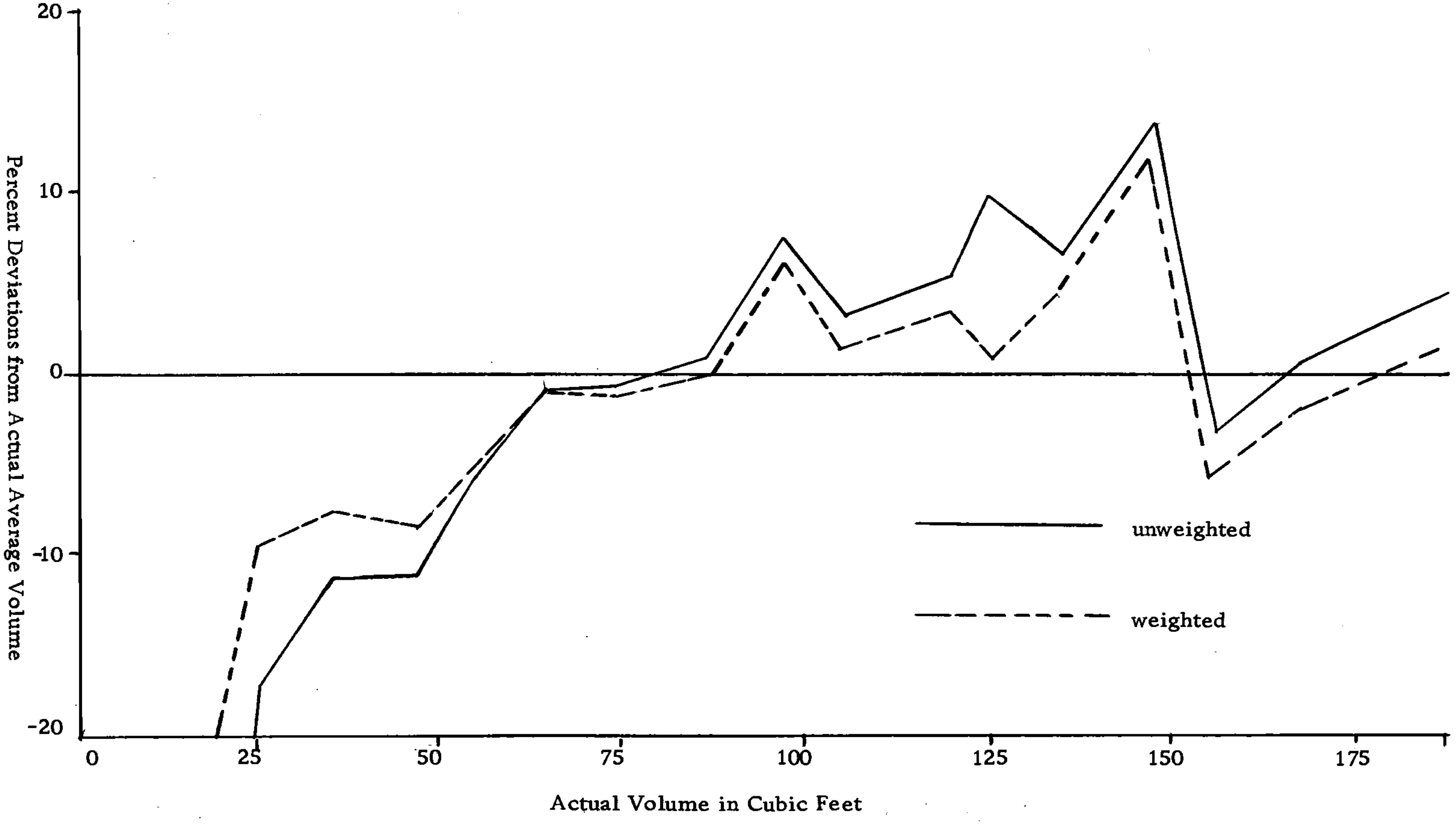
18. Steel, R. G. D. and James H. Torrie. Principles and procedures of statistics. New York, McGraw-Hill, 1960. 481 p.

19. Stoate, T. N. The use of a volume equation in pine stands. Australian Forestry 9: 48-52. 1945.

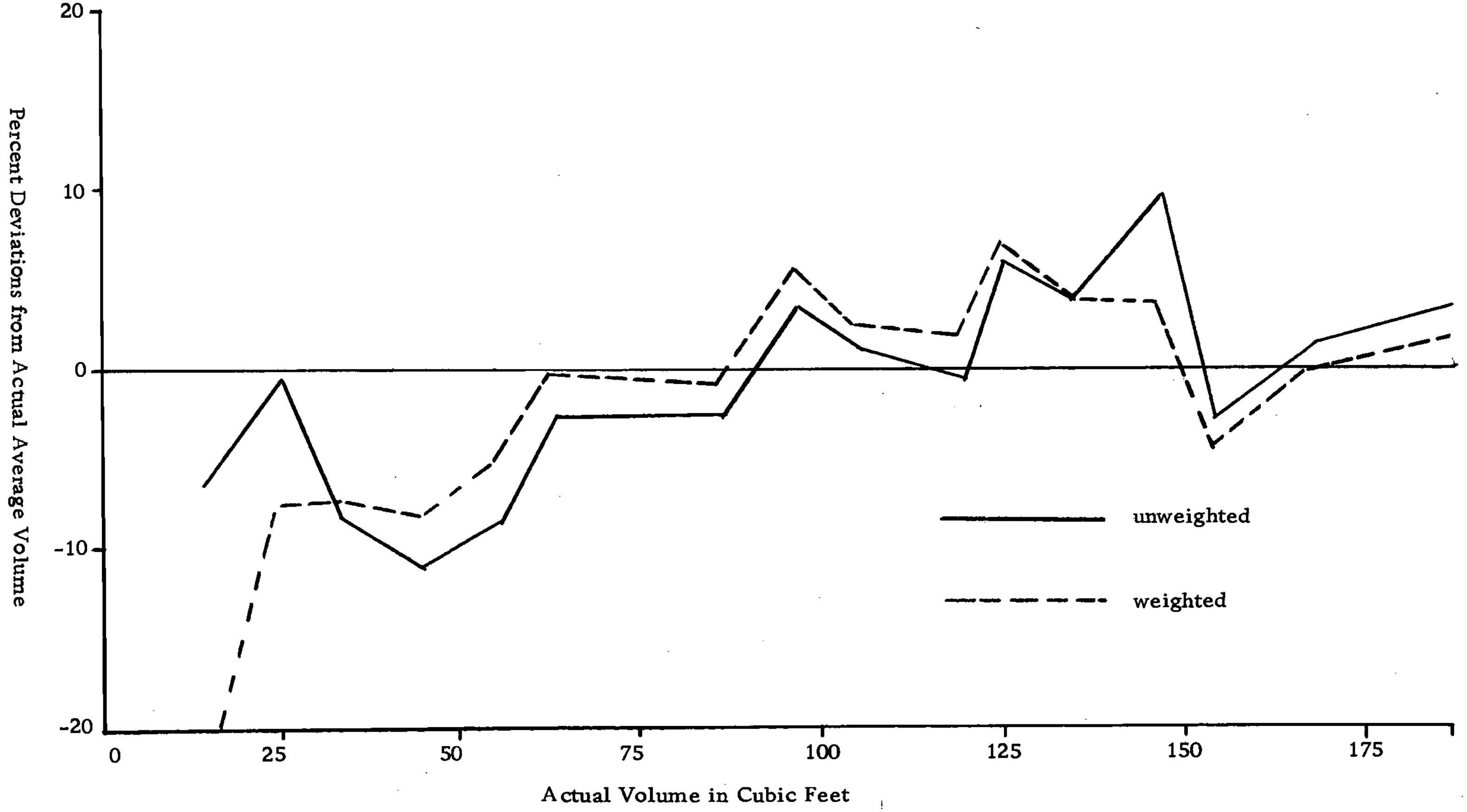
20. University of British Columbia. Forest Club. Forestry handbook for British Columbia. Vancouver, 1960. 800 p.

21. Williams, E. J. Regression analysis. New York, Wiley, 1959. 214 p.

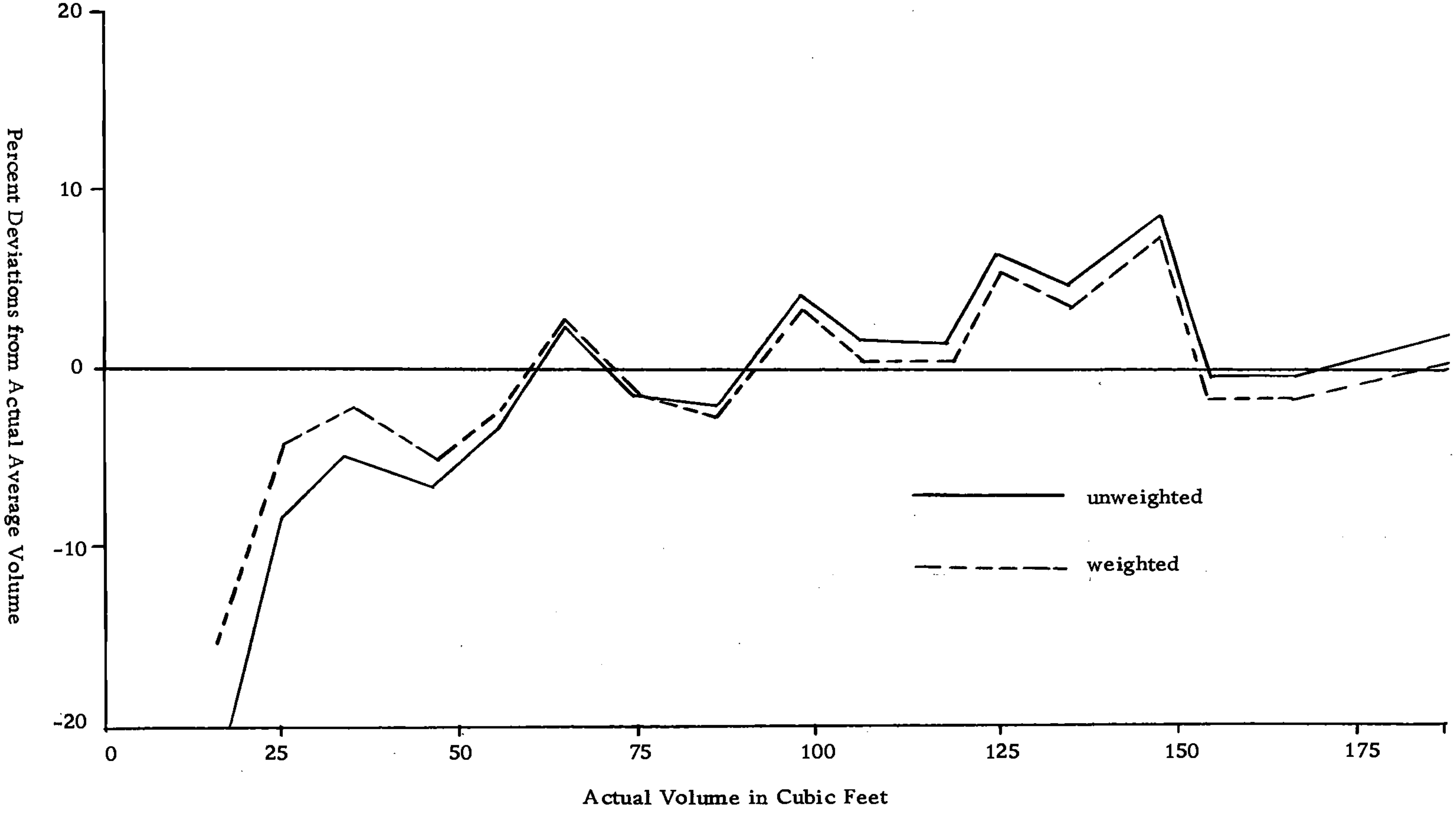
APPENDIX



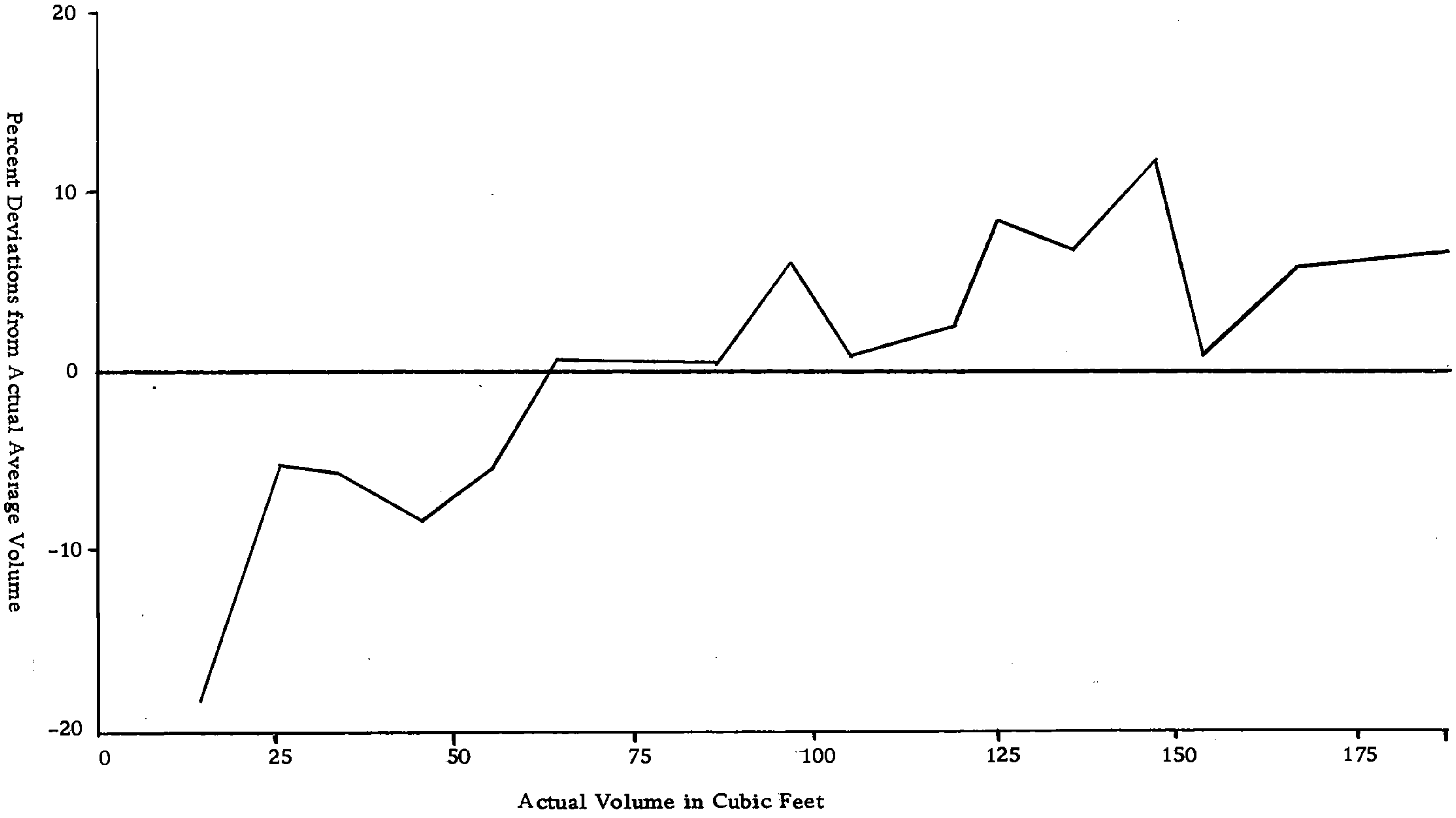
Appendix 1. Percent deviations of estimated volumes for weighted and unweighted regression solutions - combined variable equation.



Appendix 2. Percent deviations of estimated volumes for weighted and unweighted regression solutions - Australian equation.



Appendix 3. Percent deviations of estimated volumes for weighted and unweighted regression solutions - short cut form class equation.



Appendix 4. Percent deviations of estimated volumes for Schumacher's equation.

Appendix 5

Actual Tree Volumes and Estimated Tree Volumes for Test Equations

Actual Volume Cubic Feet		No. of	Estimated Volume, Cubic Feet						
			D^2H		FD^2H		Australian		
Class Interval	Average	Trees	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Schumacher
10.0 - 19.9	15.8	17	22.3	20.0	19.6	18.3	16.2	19.9	18.7
20.0 - 29.9	25.1	40	29.4	27.4	27.3	26.2	25.2	27.1	26.4
30.0 - 39.9	34.4	41	38.4	36.9	36.1	35.2	37.2	37.0	36.7
40.0 - 49.9	43.3	29	47.9	46.9	46.2	45.5	48.2	47.0	47.0
50.0 - 59.9	55.2	33	58.5	58.0	56.9	56.6	60.1	58.2	58.1
60.0 - 69.9	64.4	25	65.3	65.1	62,8	62.6	65.9	64.4	64.2
70.0 - 79.9	74.7	37	75.4	75.7	75.8	75.9	76.5	75.0	74.3
80.0 - 89.9	86.1	15	85.8	86.7	87.9	88.3	88.4	86.8	85.9
90.0 - 99.9	94.5	13	87.8	88.7	90.6	91.1	91.3	89.3	88.3
100 - 109	105.0	16	101.8	103.5	103.5	104.3	103.8	102.5	100.4
110 - 119	117.5	4	111.4	113.5	115.9	117.0	117.4	115.4	114.6
120 - 129	125.1	14	112.8	115.0	117.1	118.2	117.6	116.3	114.8
130 - 139	134.9	10	126.2	129.1	128.7	130.1	129.6	129.5	126.4
140 - 149	143.1	6	123.2	125.9	130.9	132.4	129.1	127.8	126.0
150 - 159	153.3	5	158.1	162.5	154.1	156.2	157.8	160.3	152.1
160 - 169	165.2	6	164.3	169.0	165.9	168.2	163.1	165.2	155.4
170 - 209	188.2	11	180.5	186.1	184.7	187.6	181.5	185.4	175.9
210 - 249	232.7	6	224.2	232.0	225.0	228.8	222.0	230.7	214.1
250 - 289	274.6	5	271.1	281.2	267.4	272.3	268.5	282.6	260.7
290 - 329	293.6	4	288.6	299.5	305.1	311.0	285.7	302.2	279.4
330 - +	414.2	3	396.8	413.1	411.6	420.1	385.9	414.8	369.6

Appendix 6

Volume Table for Douglas Fir*

dbh	Total Height in Feet																			No. of Trees
	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190	
2	2.7	2.7	2.8	2.9																
4	2.9	3.1	3.4	3.6																
6	3.2	3.8	4.3	4.9	5.5															
8	3.6	4.7	5.7	6.7	7.7	8.8														
10	4.2	5.8	7.4	9.0	11	12	14	15	17	19	20									3
12	4.9	7.2	9.5	12	14	16	19	21	23	26	28	30	33							18
14		8.9	12	15	18	21	25	28	31	34	37	40	43	47						31
16		11	15	19	23	27	31	35	40	44	48	52	56	60						42
18		13	18	23	29	34	39	44	49	54	60	65	69	75	80					41
20			22	28	35	41	47	54	60	67	73	79	86	92	99					47
22			26	34	41	49	57	65	72	80	88	96	103	111	119	127				48
24			30	40	49	58	67	76	86	95	104	113	127	132	141	150				32
26				46	57	68	78	89	100	111	122	132	143	154	165	176	186			20
28				53	65	78	90	103	116	128	141	153	166	178	191	203	216			18
30				60	75	89	103	118	132	147	161	175	190	204	219	233	248	262		10
32					85	101	117	134	150	167	183	199	216	232	249	265	281	297		8
34						114	132	151	169	188	206	225	243	262	280	299	317	336	354	8
36						127	148	169	189	210	331	252	272	293	314	335	355	376	397	6
38							164	188	211	234	257	280	303	326	349	373	396	419	442	3
40								207	233	259	284	310	336	361	387	412	438	464	489	2
42								229	257	285	313	342	370	398	426	454	483	511	539	2
44								251	282	313	344	375	406	437	468	499	530	560	592	0
46									308	341	375	409	443	477	511	545	579	612	646	1

*Table shows volume in cubic feet of entire stem inside bark, including stump and top, without allowance for breakage, defect or trim for Girard form class .72 . Heavy lines indicate extent of basic data. Equation used: $V = 2.622715 + 0.22240\ FD^2H \cdot 10^{-6}$.